

SKRIPSI

**APAKAH KEMUNCULAN *ABERRANT RESPONSE*
MEMPENGARUHI KEMUNCULAN *OUTLIER*? : SEBUAH
STUDI SIMULASI**



Disusun Oleh:

Diyandaru Adhitya

NIM. 155120307111078

**JURUSAN PSIKOLOGI
FAKULTAS ILMU SOSIAL DAN ILMU POLITIK
UNIVERSITAS BRAWIJAYA**

2018

LEMBAR PENGESAHAN

**Apakah Kemunculan *Aberrant Response* mempengaruhi Kemunculan
Outlier? : Sebuah Studi Simulasi**

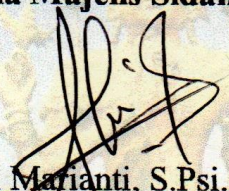
SKRIPSI

Disusun Oleh :

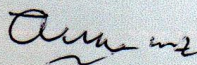
Diyandaru Adhitya
NIM. 155120307111078

Telah disetujui dan dinyatakan lulus dalam ujian sarjana
Pada tanggal **8 Oktober 2018**

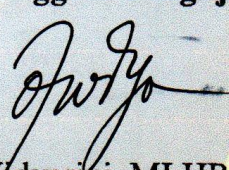
Tim Penguji
Ketua Majelis Sidang Penguji,


Sukaesi Marianti, S.Psi., M.Si., Ph.D.
NIP. 198007312008122002

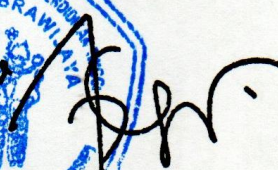
Ketua Penguji,


Nur Hasanah, S.Psi., M.Si.
NIP. 197404022008012012

Anggota Penguji,


Dr. Ika Widyarni, MLHR, Psikolog
NIK. 2007106810042001

Malang,
Fakultas Ilmu Sosial dan Ilmu Politik
Dekan


Prof. Dr. Unti Ludigdo, Ak
NIP. 19690814 199402 1 001

LEMBAR PERNYATAAN ORISINALITAS

Nama: Diyandaru Adhitya

NIM: 155120307111078

Menyatakan dengan kesungguhan bahwa skripsi dengan judul **“Apakah Kemunculan Aberrant Response mempengaruhi Kemunculan Outlier? : Sebuah Studi Simulasi”** adalah benar-benar karya saya sendiri. Hal yang bukan karya saya dalam skripsi tersebut diberi tanda kutipan dan ditunjukkan dalam daftar pustaka.

Apabila di kemudian hari terbukti bahwa pernyataan saya tidak benar, maka saya bersedia menerima sanksi akademik berupa pencabutan skripsi dan gelar yang saya peroleh dari skripsi tersebut.

Malang, 9 November 2018

Yang membuat pernyataan



Diyandaru Adhitya

155120307111078

KATA PENGANTAR

Puji syukur penulis panjatkan kepada Allah SWT karena dengan kuasa-Nya penulis dapat menyelesaikan skripsi yang berjudul “Apakah Kemunculan *Aberrant Response* mempengaruhi Kemunculan *Outlier*? : Sebuah Studi Simulasi”. Skripsi ini merupakan tugas akhir yang wajib dikerjakan oleh setiap mahasiswa pada Jurusan Psikologi Fakultas Ilmu Sosial dan Ilmu Politik Universitas Brawijaya Malang.

Penulis menyadari dalam menyelesaikan laporan ini tidaklah mudah dan membutuhkan bantuan serta dukungan dari berbagai pihak. Oleh karena itu, penulis ingin mengucapkan terima kasih kepada:

1. Orang tua dan keluarga yang sudah mendukung dan memberikan doa selama penelitian maupun penyusunan skripsi.
2. Ibu Sukaesi Marianti, S.Psi., M.Si., Ph.D sebagai dosen pembimbing skripsi yang selalu memberikan ilmu dan waktunya dalam memberikan bimbingan, arahan, dan semangat kepada penulis selama penelitian maupun penyusunan skripsi.
3. Ibu Nur Hasanah, S.Psi., M.Si. dan Dr. Ika Widyarini, MLHR, Psikolog selaku dosen penguji yang telah membantu dalam menyempurnakan skripsi ini.
4. Ibu Cleoputri Al Yusainy, Ph.D, sebagai Kepala Jurusan Psikologi Universitas Brawijaya yang turut membantu penyusunan skripsi ini.

5. Untuk sahabat penulis Danang Kamal M dan Mia Wijayanti yang senantiasa memberikan semangat, doa, dan bantuan selama pelaksanaan maupun penyusunan skripsi.
6. Untuk kelompok pertemanan penulis, Remaja Masjid, Antonim Projects, Travellers Manja, Team Laboratorium, teman-teman Mipcor, dan lain-lain yang sudah menemani hari-hari perkuliahan penulis.
7. Teman-teman mahasiswa Jurusan Psikologi Universitas Brawijaya angkatan 2015 yang telah banyak membantu penulis dan memberikan dukungan dalam penelitian maupun penyusunan skripsi.
8. Seluruh pihak yang sudah membantu penulis dalam penyelesaian penelitian maupun penyusunan skripsi ini yang tidak dapat penulis sebut satu persatu.

Akhir kata, penulis berharap Tuhan Yang Maha Esa dapat membalas segala kebaikan dan bantuan dari semua pihak. Skripsi ini tentunya tidak lepas dari kekurangan karena keterbatasan waktu, tenaga, dan pengetahuan penulis. Oleh karena itu, saran dan kritik terhadap skripsi ini sangat diperlukan untuk menyempurnakan skripsi ini. Semoga skripsi ini dapat memberikan manfaat bagi pengembangan ilmu.

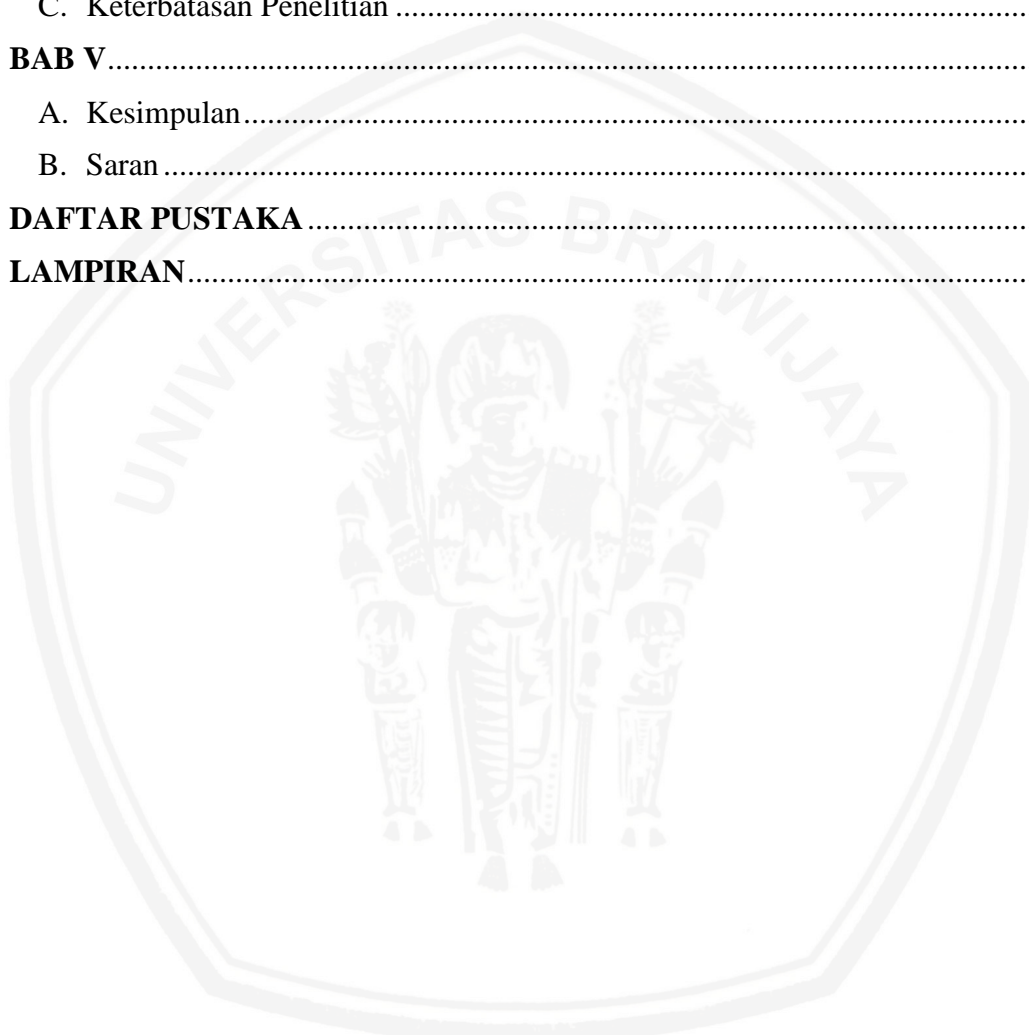
Malang, November 2018

Penulis

DAFTAR ISI

HALAMAN SAMPUL	i
LEMBAR PENGESAHAN	ii
LEMBAR PERNYATAAN ORISINALITAS	iii
KATA PENGANTAR	iv
DAFTAR ISI	vi
DAFTAR TABEL	viii
DAFTAR GAMBAR	ix
DAFTAR LAMPIRAN	x
DAFTAR ISTILAH	xi
ABSTRAK	xiii
BAB I	1
A. Latar Belakang.....	1
B. Rumusan Masalah	5
C. Tujuan Penelitian.....	5
D. Manfaat Penelitian.....	5
E. Penelitian Terdahulu.....	6
BAB II	9
A. <i>Person-fit</i>	9
B. <i>Outlier</i>	15
C. Studi Simulasi Monte Carlo	18
D. Kerangka Pemikiran	19
BAB III	20
A. Desain Penelitian	20
B. Fokus Penelitian	21
C. Definisi Operasional.....	21
D. Penetapan Populasi dan Sampel	22
E. Tahap Pelaksanaan Penelitian	23

F. Teknik Analisis Data	25
BAB IV	28
A. Hasil Penelitian.....	28
B. Pembahasan	39
C. Keterbatasan Penelitian	43
BAB V	44
A. Kesimpulan.....	44
B. Saran	44
DAFTAR PUSTAKA	46
LAMPIRAN.....	50



DAFTAR TABEL

Tabel 1. <i>Statistik-statistik person-fit kategori parametrik dan non-parametrik</i>	10
Tabel 2. <i>Nilai true parameter dan estimated parameter dari data simulasi</i>	28
Tabel 3. <i>Nilai bias dan RMSE dari tiap-tiap kondisi</i>	29
Tabel 4. <i>Parameter simulasi respon yang ekstrim</i>	30
Tabel 5. <i>Hasil uji person-fit pada data set</i>	31
Tabel 6. <i>Hasil penghitungan modified z-score pada data set</i>	31
Tabel 7. <i>Nilai item discrimination dan item location dari tiap item real data</i>	35
Tabel 8. <i>Informasi nilai lz dari real data</i>	37
Tabel 9. <i>Informasi nilai modified z-score dari real data</i>	38



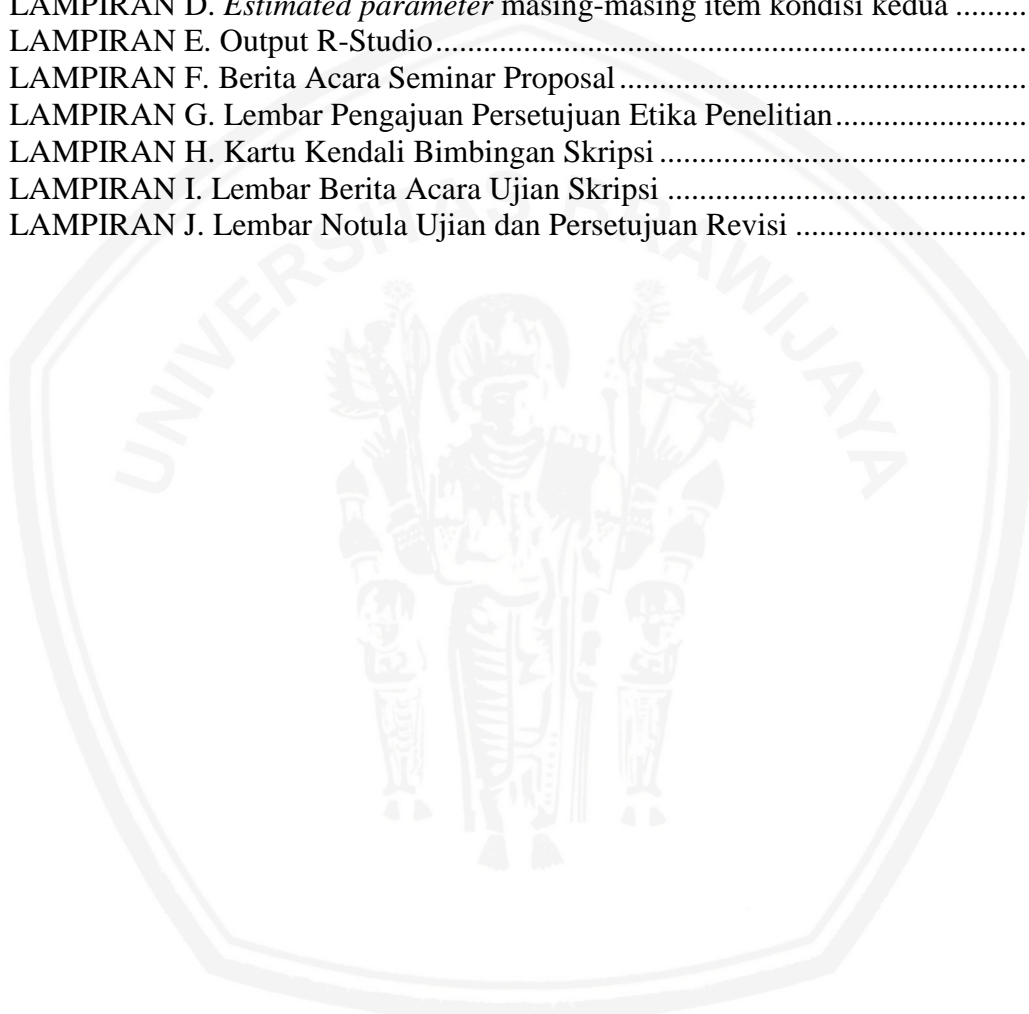
DAFTAR GAMBAR

<i>Gambar 1.</i> Tahapan penelitian menggunakan studi simulasi dan real data	20
<i>Gambar 2.</i> Hasil penghitungan nilai l_z dari masing-masing kondisi item	32
<i>Gambar 3.</i> Nilai M_i (<i>Modified z-score</i>) dari masing-masing kondisi item	33
<i>Gambar 4.</i> Hasil korelasi dari masing-masing kondisi item	34
<i>Gambar 5.</i> Hasil penghitungan l_z pada real data	37
<i>Gambar 6.</i> Hasil penghitungan nilai M_i (<i>modified z-score</i>) pada real data	38
<i>Gambar 7.</i> Hasil korelasi dari real data	39



DAFTAR LAMPIRAN

LAMPIRAN A. <i>True parameter</i> masing-masing item kondisi pertama.....	50
LAMPIRAN B. <i>Estimated parameter</i> masing-masing item kondisi pertama	51
LAMPIRAN C. <i>True parameter</i> masing-masing item kondisi kedua	52
LAMPIRAN D. <i>Estimated parameter</i> masing-masing item kondisi kedua	54
LAMPIRAN E. Output R-Studio.....	56
LAMPIRAN F. Berita Acara Seminar Proposal	59
LAMPIRAN G. Lembar Pengajuan Persetujuan Etika Penelitian.....	60
LAMPIRAN H. Kartu Kendali Bimbingan Skripsi	61
LAMPIRAN I. Lembar Berita Acara Ujian Skripsi	62
LAMPIRAN J. Lembar Notula Ujian dan Persetujuan Revisi	63



DAFTAR ISTILAH

Aberrant response Pola respon jawaban yang dianggap tidak mungkin terjadi sehingga menghasilkan nilai yang terlalu tinggi atau terlalu rendah.

Bias Rata-rata dari perbedaan antara nilai yang diestimasi dan nilai asli

Estimated parameter Parameter yang dihitung setelah data digenerasi atau setelah mendapatkan *real data* dengan menggunakan model IRT.

Mahalanobis Distance Salah satu cara untuk menghitung jarak antara poin dan distribusi dalam ruang multivariat.

Model L_z Salah satu model statistik dari *person-fit*, merupakan standar estimasi dari *lo*.

Modified z-score Bentuk modifikasi dari z-score biasa, dimana dalam penghitungannya menggunakan median dan *Median Absolute Deviation* (MAD)

Outlier Istilah dalam statistik yang menggambarkan data yang berbeda secara ektrim dari keseluruhan data set.

Parameter recovery Tahapan dalam studi simulasi untuk mengetahui kelayakan data dengan membandingkan *true parameter* dan *estimated parameter*.

Pearson's product moment Uji pengukuran korelasi linear antara variabel X dan variabel Y.

Person-fit Sekumpulan model statistik yang memiliki fungsi untuk mendeteksi pola respon yang menyimpang.

Polytomous data Data set yang memiliki item dengan pilihan skor jawaban lebih dari dua.

RMSE (Root Mean Square Deviation) Pengukuran yang digunakan untuk melihat perbedaan antar nilai, pada penelitian ini untuk mengukur perbedaan *true parameter* dengan *estimated parameter*.

True Parameter Parameter asli yang ditentukan oleh peneliti saat menggenerasi data tanpa terpengaruh oleh *error*.

APAKAH KEMUNCULAN *ABERRANT RESPONSE* MEMPENGARUHI KEMUNCULAN *OUTLIER*? : SEBUAH STUDI SIMULASI

Diyandaru Adhitya, Sukaesi Marianti
adhit716@gmail.com

Jurusan Psikologi, Universitas Brawijaya, Malang, Indonesia

ABSTRAK

Penelitian ini memiliki tujuan untuk mengetahui kemunculan *aberrant response* sebagai penyebab dari kemunculan *outlier*. Metode penelitian yang digunakan adalah studi simulasi dan pengaplikasian pada *real data*. Pada studi simulasi digunakan 2 kondisi jumlah item, yaitu 20 item dan 40 item dengan 6 pilihan jawaban. Masing-masing kondisi terdiri dari 450 responden yang berasal dari distribusi normal yang kemudian dikontaminasi 50 responden dengan respon ekstrim. Data kemudian diujikan dengan menggunakan *person-fit* dan *modified z-score*. Hasil kedua uji tersebut kemudian dikorelasikan dengan menggunakan *Pearson's product moment*. Hasil dari penelitian ini menunjukkan bahwa tidak ada korelasi ($r = 0.048$, $r = -0.044$, dan $r = 0.095$) antara kemunculan *aberrant response* terhadap kemunculan *outlier*.

Keywords: person-fit, outlier, aberrant response, studi simulasi.

DOES THE APPEARANCE OF ABERRANT RESPONSE AFFECT THE APPEARANCE OF OUTLIERS? : A SIMULATION STUDY

Diyandaru Adhitya, Sukaesi Marianti
adhit716@gmail.com

Psychology Major, Brawijaya University, Malang, Indonesia

ABSTRACT

This study aims to determine the appearance of aberrant response as the cause of the appearance of outliers. The method that were used in this research are simulation study and the application on real data. In the simulation study two conditions were used, 20 items as the first condition and 40 items as the second with 6 answer choices. Each condition consists of 450 respondents that were taken from normal distribution and then contaminated by 50 respondents with extreme responses. The data is then tested using person-fit and modified z-score. The results of the two tests were then correlated using Pearson's product moment. The results of this study indicate that there are no correlations ($r = 0.048$, $r = -0.044$, and $r = 0.095$) between the appearance of aberrant response and the appearance of outliers.

Keywords: person-fit, outlier, aberrant response, simulation study.

BAB 1

PENDAHULUAN

A. Latar Belakang

Tes psikologis merupakan salah satu senjata seorang Psikolog maupun Ilmuwan Psikologi dalam menentukan posisi seseorang pada sifat yang diukur (Meijer, 2003). Tes psikologis pun kerap dipergunakan dalam pengambilan keputusan di dunia pendidikan, seperti menentukan kelayakan individu untuk mengenyam pendidikan tertentu atau bahkan menerima gelar pendidikan tertentu (Karabatsos, 2003). Oleh karena itu, suatu tes psikologis penting untuk memiliki pengukuran yang akurat, karena pengukuran yang tidak akurat dapat menyebabkan konsekuensi negatif bagi penerima maupun pemberi tes. Skor tidak akurat yang terlalu tinggi dapat menyebabkan individu terdaftar dalam program pendidikan atau mendapat gelar pendidikan tertentu walaupun ia tidak memenuhi syarat. Sebaliknya, individu dengan skor tidak akurat yang terlalu rendah dapat dikeluarkan atau ditolak secara tidak adil dari program akademik walaupun ia memenuhi syarat.

Adanya kepentingan keakuratan tes psikologis tersebut menyebabkan mulainya dilakukan studi untuk mendeteksi pola skor item yang tidak mungkin (*aberrant*) pada tahun 1981 (Meijer, 1996). Pada ilmu Psikometrika, skor item

aberrant muncul ketika responden mendapatkan jawaban yang benar atau salah namun didasari alasan yang salah. Contohnya adalah responden A mendapatkan jawaban yang benar dengan alasan ia melihat responden lain memilih jawaban tersebut. Contoh lainnya adalah responden B yang mendapat jawaban yang salah karena ia belum terbiasa menggunakan jenis alat tes tersebut. Menurut Meijer dan Sijsma (2001), terdapat 3 sebab yang menyebabkan timbulnya *aberrant response*. Pertama adalah “*sleeping*” *behavior*, dimana responden memberikan respon yang salah pada pertanyaan mudah di awal tes karena belum beradaptasi dengan tes. Kedua adalah *cheating*, dimana responden menyamakan jawaban dengan responden lain. Terakhir adalah “*plodding*” *behavior*, dimana responden bekerja sangat lambat dan cermat, sehingga menghasilkan pola skor-item yang terlalu ideal. Namun, tidak semua jenis perilaku yang tidak biasa mempengaruhi skor tes. Munculnya respon item yang *aberrant* tergantung pada banyak faktor, seperti jenis dan jumlah perilaku yang tidak biasa.

Adanya konsekuensi yang dihasilkan dari *aberrant response* membuatnya sangat penting untuk diidentifikasi pada suatu alat tes. Statistik *person-fit* dirancang untuk menemukan responden dengan pola respon item yang *aberrant* (yang mengarah ke skor tes terlalu tinggi atau rendah), dan juga untuk membedakan responden tersebut dari responden dengan pola respon-item yang normal (Karabatsos, 2003). Lebih mudahnya, statistik *person-fit* mengukur tingkat kewajaran jawaban responden terhadap serangkaian item tes. Dalam tinjauan

metodologi yang dilakukan oleh Meijer dan Sijsma (2001), terdapat 36 statistik *person-fit* yang berfungsi untuk mengidentifikasi responden dengan *aberrant response*. Menurut Meijer, Niessen, dan Tendeiro (2016) untuk item yang bersifat dikotomi dan menggunakan analisis parametrik IRT, direkomendasikan statistik l_z^* karena ketersediaan skor *cutoff* yang mudah diinterpretasi. Sedangkan untuk item yang bersifat *polytomous* dan tes yang panjang, statistik l_z merupakan yang paling sering digunakan karena interpretasinya yang mudah. Salah satu karakteristik utama dari *aberrant response* adalah ketidaksesuaian dengan model pengukuran yang digunakan, hal ini hampir serupa dengan *outlier*, dimana mereka mewakili setiap variabel yang nilainya tidak sesuai dengan model statistik yang digunakan (Widhiarso & Sumintono, 2016).

Outlier didefinisikan sebagai nilai variabel yang sangat tidak biasa atau mustahil, dibandingkan dengan distribusi semua nilai yang valid (Bethlehem, 2009). Menurut Lee (Widhiarso & Sumintono, 2016) munculnya *outlier* sering tidak dapat dihindari ketika melakukan survei dengan menggunakan kuesioner, terutama ketika data dikumpulkan untuk berbagai tujuan. Apabila *aberrant response* disebabkan oleh perilaku dari responden yang mengisi tes, data *outlier* disebabkan oleh dua kategori, *outlier* yang timbul dari kesalahan dalam data, dan *outlier* yang timbul dari variabilitas yang melekat pada data (Anscombe, 1960). Sangat penting bagi peneliti untuk mengidentifikasi *outlier*, karena mereka mungkin dapat membuat data statistik menjadi bias dan menyebabkan interpretasi

yang salah. Outlier dapat dideteksi pada data set berjenis *univariate* maupun *multivariate*, namun dengan model statistik yang berbeda. Data set berjenis *univariate* dianjurkan untuk menggunakan *modified Z-score* sebagai model statistik (Iglewicz & Hoaglin, 1993). Sedangkan untuk data set berjenis *multivariate* dianjurkan untuk menggunakan *Mahalanobis Distance* sebagai model statistik (Penny, 1996).

Karena kemiripan karakteristik dari *aberrant response* dan *outlier*, Widhiarso dan Sumintono (2016) melakukan penelitian dengan tujuan untuk menyelidiki korelasi antara kemunculan *aberrant response* dengan kemunculan *outlier* pada data. Namun, berdasarkan analisis dari hasil penelitiannya didapatkan bahwa responden yang secara stabil menghasilkan *aberrant response*, mereka tidak selalu dianggap sebagai *outlier*. Hal ini mungkin disebabkan karena ukuran sampel yang terlalu besar ($N = 2983$) yang dapat meminimalkan kemunculan dari *outlier*. Oleh karena itu, penelitian kali ini bertujuan untuk menguji kembali korelasi antara kemunculan *aberrant response* dengan kemunculan *outlier* pada data, namun dengan responden yang lebih sedikit dan menggunakan data simulasi. Data simulasi digunakan dengan alasan karena data empiris (*real data*) sering tidak lengkap, dimana hal ini dapat mempengaruhi hasil penelitian, terutama ketika jumlah data yang hilang cukup besar dan pola kehilangan bersifat tidak random (Bulut & Sunbul, 2017). Selain itu, penelitian ini juga menggunakan *real data* dengan jumlah responden yang sama agar dapat dibandingkan dengan data

simulasi. Dengan meneliti korelasi antara kemunculan *aberrant response* dengan kemunculan *outlier* pada data, dapat bermanfaat bagi peneliti agar lebih akurat dalam mengidentifikasi karakteristik individu mana yang paling mungkin terdeteksi sebagai *outlier* dan peneliti juga dapat mengurangi jumlah peserta yang dianggap *outlier* secara keliru sehingga tidak perlu mengeluarkannya dari analisis.

B. Rumusan Masalah

Berdasarkan latar belakang, maka adapun perumusan masalah dari penelitian ini adalah apakah *aberrant response* yang diidentifikasi dengan *person-fit* juga akan teridentifikasi sebagai *outlier*?

C. Tujuan Penelitian

Berdasarkan perumusan masalah, maka adapun tujuan dari penelitian ini adalah untuk menguji korelasi kemunculan *aberrant response* (diidentifikasi dengan *person-fit*) dengan kemunculan *outlier*.

D. Manfaat Penelitian

1. Manfaat teoretis

Secara teoretis, penelitian ini diharapkan dapat memberikan sumbangan terhadap ilmu pengetahuan, khususnya dalam bidang ilmu psikologi dan psikometrika yang berkaitan dengan Uji *Person-fit*, Studi Simulasi dan *Outlier*.

Selain itu, penelitian ini diharapkan dapat memberikan sumbangan dan masukan dalam pengembangan skala tes psikologis di Indonesia.

2. Manfaat praktis

Hasil penelitian ini dapat menjadi pertimbangan bagi peneliti mengenai *aberrant response* sebagai alasan penghapusan data-data *outlier* yang mengganggu hasil akhir penelitian.

Selain itu, diharapkan juga hasil penelitian ini dapat menjadi rujukan dan pembanding bagi penelitian berikutnya yang relevan.

E. Penelitian Terdahulu

Pada perencanaan dan pelaksanaan penelitian ini, peneliti menggunakan beberapa penelitian terdahulu sebagai ahan rujukan, berikut adalah beberapa penelitian yang dirujuk:

1. Felt, Castaneda, Tiemensma, dan Depaoli pada tahun 2017 melakukan penelitian mengenai penggunaan statistik person-fit untuk mendeteksi outliers pada riset survey. Tujuan dari penelitian ini adalah untuk menggambarkan bagaimana memperoleh statistik person-fit sebagai pendeteksi outlier melalui metode statistik yang memeriksa kuesioner berbasis kesehatan. Penelitian ini menggunakan pasien yang menderita *Cushing Syndrome* ($n = 394$), yang diarahkan untuk mengerjakan suatu survey online berisikan *CushingQoL* (*Cushing Quality of Life*). Metode deteksi outlier yang konvensional (*boxplots*)

mengungkapkan tidak ada outlier yang mencerminkan skor ekstrim pada hasil CushingQoL. Namun, statistik person-fit (Statistik Zh) mengidentifikasi 18 pasien dengan pola respon "atypical".

2. Karabatsos pada tahun 2003 meneliti kemampuan statistic *person-fit* sebagai pendeteksi *aberrant response*. Penelitian ini memiliki tujuan untuk membandingkan 36 indeks *person-fit*, di bawah kondisi pengujian yang berbeda, untuk mendapatkan konsensus yang lebih baik mengenai statistik person-fit mana yang terbaik. 36 statistik person-fit dibandingkan dalam konteks model *Rasch*. Desain penelitian termasuk jenis-jenis respon partisipan yang atipikal, persentase munculnya respon partisipan yang atipikal, dan jenis panjang tes. Dari seluruh 36 statistik person-fit, didapatkan bahwa statistik H^T , $D(\theta)$, C, MCI, dan U3 merupakan yang terbaik dalam mendeteksi respon atipikal partisipan. Namun secara keseluruhan, statistik H^T merupakan yang terbaik dalam mengidentifikasi respon tes yang atipikal. Statistik H^T juga merupakan salah satu yang terbaik dalam mendeteksi masing-masing dari jenis-jenis respon partisipan yang atipikal, dan dalam mendeteksi respon partisipan pada masing-masing jenis panjang tes.
3. Penelitian Artner pada tahun 2016 menggunakan studi simulasi untuk membandingkan lima indeks person-fit yang dikenal dalam hal spesifitas dan sensitivitas, di bawah kondisi pengujian yang berbeda. Terdapat 4 parameter untuk menghasilkan scenario yang berbeda. 4 parameter tersebut adalah jumlah item, jumlah responden, proporsi respon atipikal, dan jenis-jenis respon yang

atipikal. Temuan pada penelitian ini memang berbeda secara substansial dari studi simulasi lainnya. Rata-rata, H^T memiliki area terbesar di bawah kurva ROC, namun C^* dan $U3$ berkinerja lebih buruk lagi.

4. Pada tahun 2016, Widhiarso dan Sumintono melakukan penelitian mengenai aberrant response sebagai penyebab adanya outlier dalam analisis statistic. Peserta penelitian ini adalah siswa sekolah menengah atas ($N = 2983$) yang mengisi tiga kuesioner kepribadian. *Aberrant response* untuk instrumen ini dideteksi menggunakan statistik infit, outfit, dan person-fit di bawah pemodelan Rasch, yang kemudian hasilnya akan dikategorisasikan menjadi overfit, fit, dan underfit. Mahalanobis Distance (MD) digunakan untuk mengidentifikasi responden yang diklasifikasikan sebagai outlier, berdasarkan analisis regresi sederhana. Studi ini menemukan bahwa responden yang terkategori underfit lebih cenderung menunjukkan nilai MD yang lebih tinggi daripada responden overfit atau fit, yang berarti bahwa mereka cenderung diklasifikasi sebagai outlier.

BAB II

TINJAUAN PUSTAKA

A. Person-fit

1. Pengertian *person-fit*

Person-fit merupakan suatu model statistik yang berfungsi untuk mendeteksi keberadaan respon yang menyimpang dari asumsi. Berdasarkan penelitian yang dilakukan oleh Felt et al. (2017), diketahui juga uji *person-fit* lebih baik dalam mendeteksi pola respon atipikal dibandingkan metode deteksi konvensional lainnya. Menurut Meijer dan Sijtsma (2001), metode *person-fit* dipergunakan untuk mengevaluasi ketidaksesuaian kinerja tes individu pada model IRT atau pola item-skor lainnya dalam sampel berupa sekelompok orang. Ketika pola respon dianggap menyimpang, maka diasumsikan bahwa respon terhadap item survei dipandu oleh mekanisme respon selain konstruk yang ditentukan.

2. Statistik-statistik *person-fit*

Person-fit berisikan seperangkat metode statistik yang digunakan untuk mengidentifikasi pola respon menyimpang. Di dalam Meijer dan Sijtsma (2001) dijabarkan 36 statistik dalam person-fit yang telah terklasifikasi sebagai parametrik dan non-parametrik. Berikut adalah penjabarannya:

Tabel 1.

Statistik-statistik person-fit kategori parametrik dan non-parametrik

<i>Non-Parametric</i>	<i>Parametric</i>
G	U
G^*	ZU
r_{pbis}	$\ln U$
C	W
MCL	ZW
$U3$	$\ln W$
$ZU3$	$ECI1, ECI2, ECI3, ECI4, ECI5, ECI6,$
H^T	$ECI1z, ECI2z, ECI4z, ECI6z$
A, D, E_i	l
	l_z
	$.M$
	$.M(p\text{-value})$
	<i>Item-Grouping Person-fit Statistics</i>
	$D(\theta)$
	l_{zm}
	UB
	ZUB
	$\ln UB$

Menurut Meijer, Niessen, dan Tendeiro (2016), untuk item yang bersifat dikotomi dan menggunakan analisis parametrik IRT, direkomendasikan statistik l_z^* karena ketersediaan skor *cutoff* yang mudah diinterpretasi. Sedangkan untuk item yang bersifat politomi dan tes yang panjang, statistik l_z merupakan yang paling sering digunakan karena interpretasinya yang mudah. Karena penelitian ini menggunakan item yang bersifat politomi, maka akan digunakan statistik l_z .

3. Statistik l_z

Beberapa penelitian sebelumnya telah menemukan bahwa statistik l_z dapat bekerja lebih baik daripada statistik *person-fit* lainnya dalam banyak kasus. Salah satunya adalah penelitian yang dilakukan oleh Li & Olejnik (1997), dimana mereka membandingkan kinerja dari statistik l_z , ZU , ZW , $ECI2_z$ dan $ECI4_z$. Penelitian tersebut menemukan bahwa statistik l_z adalah statistik terkuat, karena ia dapat mengidentifikasi dua pertiga dari keseluruhan *aberrant response*. l_z merupakan standar estimasi dari l_o , sehingga untuk mencari nilai l_z harus terlebih dahulu mencari nilai l_o . Berikut adalah persamaan (1) untuk mencari nilai l_o .

$$l_o = \sum_{i=1}^n \{u_i \ln P_i(\theta) + (1 - u_i) \ln [1 - P_i(\theta)]\} \quad (1)$$

Dimana n merupakan jumlah item pada tes, u_i merupakan respon individu pada item ke- i , dan $P_i(\theta)$ merupakan kemungkinan respon yang diberikan pada item i dengan individu berkemampuan θ . Kemudian, dilanjutkan dengan persamaan (2) untuk mencari nilai l_z .

$$l_z = \frac{l_o - E(l_o)}{[Var(l_o)]^{1/2}} \quad (2)$$

Dimana nilai $E(l_o)$ merupakan *expected value* dari l_o dan $Var(l_o)$ merupakan nilai variansi dari l_o . Pola respon dari responden yang ditunjukkan oleh nilai negatif yang

besar dari l_z (biasanya kurang dari -2.00) maka dapat diidentifikasi sebagai pola respon yang atipikal atau tidak mungkin.

4. Pengertian *aberrant response*

Aberrant response merupakan pola respon pada item data set yang dianggap tidak mungkin (memiliki skor terlalu tinggi atau terlalu rendah). Kemunculan *aberrant response* ini disebabkan oleh perilaku dari responden tes. Sebagai contoh, seseorang dapat menebak beberapa jawaban yang benar untuk item pilihan ganda pada tes kecerdasan, sehingga meningkatkan total skornya pada tes dengan keberuntungan dan bukan oleh kemampuan. Demikian pula, seseorang yang tidak akrab dengan format tes dapat memperoleh skor yang lebih rendah dari yang diharapkan berdasarkan tingkat kemampuannya. Namun apakah perilaku menyimpang akan mengarah pada munculnya pola skor-item *aberrant* tergantung pada banyak faktor, seperti jenis dan intensitas perilaku menyimpang (Meijer & Sijtsma, 2001).

5. Sebab-sebab *aberrant response*

Kemunculan *aberrant response* pada suatu data set dapat disebabkan oleh perilaku-perilaku menyimpang responden dalam tes. Menurut Meijer dan Sijtsma (2001), terdapat 3 perilaku penyebab timbulnya *aberrant response*, yaitu:

a. Sleeping behavior

Terjadi ketika seorang partisipan mengalami kesulitan untuk memulai mengerjakan tes, dan setelah mulai terbiasa dengan tugasnya, ia tidak memeriksa jawaban pada beberapa item yang lebih mudah dalam ujian.

b. Cheating

Terjadi ketika seorang partisipan mencontoh atau menyalin hasil kerja dari partisipan lain.

c. Plodding Behavior

Terjadi ketika seorang partisipan bekerja sangat lambat dan sangat memperhatikan jawabannya sehingga menghasilkan pola respon item yang terlalu ideal.

6. Skala *Plasticity/Stability*

Pada penelitian ini menggunakan skala PS atau *Plasticity/Stability*. Skala PS disusun oleh DeYoung, dimana ia membagi jenis kepribadian ke dalam *plasticity* dan *stability*. Skala PS terdiri dari 40 item, dimana 20 item menggambarkan kepribadian *plasticity* dan 20 item menggambarkan kepribadian *stability*. Setiap item terdiri dari 6 pilihan jawaban. Skala PS dibuat berdasarkan skala *Big Five Personality*, sehingga ketika diujikan validitasnya *Plasticity* lebih banyak menggambarkan trait *Extraversion* dan *Openness*, sedangkan *Stability*

lebih banyak menggambarkan trait *Agreeableness*, *Conscientiousness*, dan *Neuroticism* (Hirsh, DeYoung, & Peterson, 2009).

Skala PS merupakan tes berjenis *low-stake test*. Cole dan Osterlind (2008) mendefinisikan *low-stakes test* sebagai suatu tes yang tidak memiliki konsekuensi yang berarti terhadap partisipan tes. Sedangkan *high-stake test* memiliki setidaknya beberapa konsekuensi akademis atau konsekuensi bermakna lainnya bagi partisipan tes. Karena perbedaan konsekuensi tersebut, maka kinerja yang ditunjukkan partisipan tes pun berbeda. Seperti pada penelitian yang dilakukan oleh Cole dan Osterlind (2008), yang memiliki kesimpulan bahwa sekelompok mahasiswa yang diminta untuk mengikuti tes tanpa konsekuensi (*low-stake tests*) memberikan performansi lebih rendah dibandingkan sekelompok mahasiswa yang mengikuti tes yang sama dengan konsekuensi lebih tinggi (*high-stake tests*).

Uji *person-fit* dapat mendeteksi *aberrant response* pada kedua jenis tes tersebut. Namun, tentu akan memiliki hasil yang berbeda karena kedua jenis tes tersebut menghasilkan partisipan dengan *self-motivation* yang berbeda (Batenburg, Hendriks, & Kuyper, 2012). Seperti pada penelitian yang dilakukan oleh Schmitt et al. (1999) dimana didapatkan bahwa terpadat korelasi positif yang kecil antara *test-taking motivation* dan hasil uji *person-fit*.

B. Outlier

1. Pengertian *outlier*

Di dalam suatu data set, sering ditemukan suatu data yang berbeda secara ekstrim dari mayoritas data set, skor tersebut disebut sebagai *outlier*. Anusha & Rao (2018) mendefinisikan outlier sebagai data dalam informasi yang tidak sesuai dengan perilaku atau informasi yang tidak sesuai dengan perilaku yang diharapkan. Diketahui *outlier* mungkin dapat membuat data statistik deskriptif dan inferensial menjadi bias dan menyebabkan interpretasi yang salah. Contohnya adalah ketika diadakan suatu tes, rata-rata responden mendapatkan nilai 10. Kemudian terdapat seorang responden yang mendapatkan skor 80. Skor responden tersebut mungkin tidak akan mempengaruhi median dari data, namun akan jelas mempengaruhi mean data.

2. *Modified Z-score*

Penelitian ini menggunakan data set yang bersifat *univariate*, sehingga digunakan *Modified Z-score* sebagai pendeteksi outlier sesuai anjuran Iglewicz dan Hoaglin (1993). Penggunaan *Z-score* berjenis *modified* dianjurkan karena dua estimator yang digunakan dalam *Z-Score* biasa, yaitu mean sampel dan standar deviasi sampel, dapat dipengaruhi oleh beberapa nilai ekstrim atau bahkan hanya dengan satu nilai ekstrim (Seo, 2006). Untuk menghindari masalah tersebut, median dan median dari deviasi mutlak median (MAD) digunakan dalam *modified*

Z-Score. Berikut adalah persamaan (3) yang digunakan untuk mendapatkan nilai *Modified Z-score* dari data.

$$M_i = \frac{0.6745(x_i - \tilde{x})}{MAD} \quad (3)$$

Dimana nilai \tilde{x} merupakan nilai median dari data dan MAD merupakan penyimpangan dari median yang absolut. Persamaan (4) berikut digunakan untuk menentukan nilai MAD dari data.

$$MAD = \text{median}(|x_i - \tilde{x}|) \quad (4)$$

Iglewicz and Hoaglin (1993) menyarankan bahwa nilai M_i yang lebih besar daripada 3.5 untuk ditandai berpotensi sebagai *outlier*.

3. Sebab-sebab *outlier*

Kemunculan data outlier pada data set dapat disebabkan oleh banyak hal. Menurut Osborne & Overbay (2004) terdapat 5 sebab kemunculan data outlier:

a. Outlier dari data yang error

Outlier sering disebabkan oleh kesalahan manusia, seperti kesalahan dalam pengumpulan data, perekaman, atau entri data. Data dari assesmen dapat dicatat dengan tidak benar, atau salah ketik pada entri data.

b. Outlier dari *misreporting* yang disengaja

Hal ini dapat terjadi ketika para responden dengan sengaja melaporkan data yang salah kepada peneliti atau surveyor. Motifnya dapat bermacam-macam, seperti sengaja untuk mensabotase penelitian (Huck dalam Osborne & Overbay, 2004), ingin memiliki presentasi diri yang baik, dan lain-lain.

c. Outlier dari kesalahan pengambilan sampel

Ada kemungkinan bahwa beberapa anggota sampel secara tidak sengaja ditarik dari populasi yang berbeda dari sisa sampel sehingga menghasilkan data yang berbeda pula dari sisa sampel.

d. Outlier dari kegagalan standardisasi

Outlier dapat disebabkan oleh metodologi penelitian, terutama jika sesuatu yang tidak wajar terjadi selama assesmen. Contohnya adalah ketika tes kognitif diadakan pada saat di luar tempat tes sedang konstruksi bangunan yang bising dan membuat para responden tes tidak fokus saat mengerjakan tes.

e. Outlier dari asumsi distribusi yang salah

Asumsi yang salah tentang distribusi data juga dapat menyebabkan keberadaan outlier yang dicurigai. Contohnya adalah seorang peneliti mengasumsikan responden-responden tesnya akan mendapatkan skor sekitar 50-80, maka selain skor tersebut akan dicurigai sebagai outlier.

C. Studi Simulasi Monte Carlo

Simulasi Monte Carlo pertama kali ditemukan oleh S. Ulam pada tahun 1944 sebagai penelitian untuk membuat bom atom pertama (Pillana, 2000). Dalam psikometrika, Simulasi Monte Carlo juga menjadi elemen kunci dari riset operasional maupun akademik. Dewasa ini, peneliti lebih memilih untuk mensimulasikan data dibandingkan mengumpulkan data empiris. Hal ini terjadi karena hampir tidak mungkin untuk menghilangkan efek dari variabel pengganggu yang terkait dengan peserta ujian (misalnya, jenis kelamin, sikap, dan motivasi) dan item tes (misalnya, konten, kompleksitas linguistik, dan kompleksitas kognitif) (Bulut & Sunbul, 2017).

Dalam simulasi, data dibuat oleh peneliti berdasarkan suatu model. Model ini dapat berbentuk apa pun, baik sesederhana distribusi normal dengan mean dan standar deviasi yang diberikan, maupun serumit respons yang dihasilkan berdasarkan model IRT multidimensional (Feinberg & Rubright, 2016).

D. Kerangka pemikiran

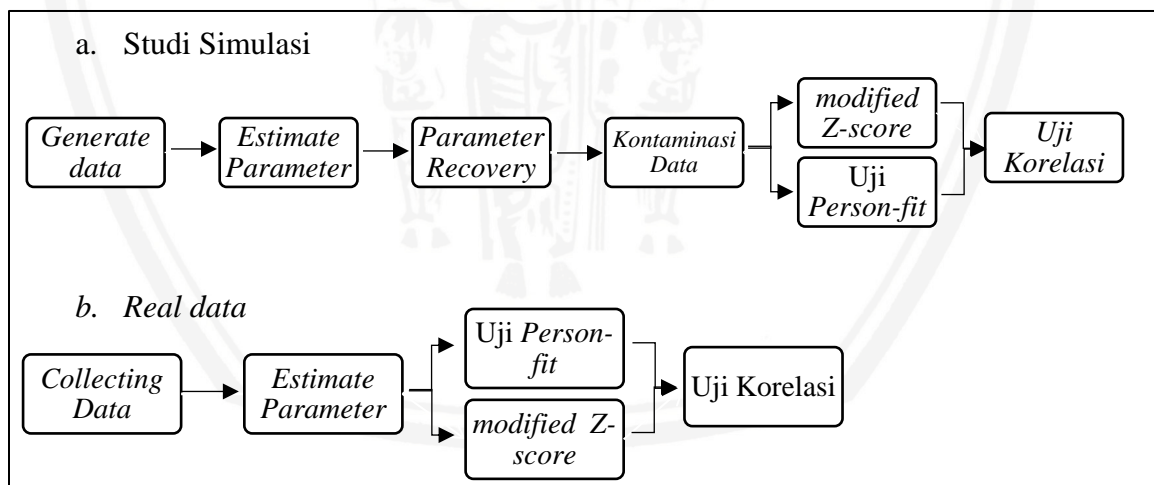
Aberrant response merupakan suatu pola respon yang menyimpang dari asumsi dan disebabkan oleh perilaku responden ketika mengerjakan tes. *Aberrant response* dapat dideteksi dengan menggunakan *person-fit*. *Outlier* merupakan data statistik yang dianggap tidak biasa atau ekstrim ketika dibandingkan dengan set datanya. Perbedaan dari aberrant response dan outlier dapat dilihat dari jenis data yang dideteksi. Aberrant response dideteksi dengan menggunakan person-fit pada kumpulan respon atau pola responnya. Sedangkan outlier dideteksi dengan menggunakan modified z-score pada hasil akhir dari data. Salah satu penyebab dari kemunculan *outlier* dalam data set adalah *human error* ketika pengambilan data atau bahkan *misreport* secara sengaja, dimana perilaku tersebut pun yang turut menghasilkan *aberrant response*. Dengan melakukan studi simulasi dan pada real data, penelitian ini bertujuan untuk menguji kemunculan *aberrant response* sebagai salah satu penyebab dari kemunculan dari *outlier*.

BAB III

METODE PENELITIAN

A. Desain Penelitian

Penelitian ini merupakan penelitian dengan pendekatan kuantitatif dan menggunakan studi simulasi. Studi simulasi yang dilakukan merupakan pendekatan IRT dan hasil dari studi simulasi diaplikasikan pada data asli (*real data*). Teknik analisis yang digunakan adalah *person-fit* untuk mendeteksi *aberrant response* dan *modified Z-score* untuk mendeteksi keberadaan *outlier*. Tahapan penelitian dapat dilihat pada gambar 1.



Gambar 1. Tahapan penelitian menggunakan studi simulasi (a) dan *real data* (b)

B. Fokus Penelitian

Fokus dalam penelitian ini adalah untuk menguji apabila terdapat korelasi antara *aberrant response* yang diidentifikasi menggunakan *person-fit* dengan kemunculan *outlier* pada data simulasi maupun data asli. Pengujian dilakukan dengan mensimulasikan data responden, dimana sebagian respon akan dibuat *aberrant*. Data simulasi kemudian akan diidentifikasi dengan *person-fit* dan dibandingkan dengan kemunculan *outlier* yang diidentifikasi dengan *modified Z-score*. Hasil dari studi simulasi diujikan kembali pada data asli.

C. Definisi Operasional

1. *Person-fit*

Person-fit merupakan suatu metode statistik yang dipergunakan untuk mendeteksi adanya respon atipikal atau *aberrant* dalam suatu set data dengan cara membandingkan data respon yang ada dengan data respon asumsi. Dalam pengukurannya, penelitian ini menggunakan statistic l_z karena lebih mudah untuk diinterpretasi dan terbukti dapat bekerja lebih baik berdasarkan beberapa studi. Pencarian nilai l_z dapat dilakukan dengan persamaan (1) dan kemudian dilanjutkan dengan persamaan (2). Pola respon dari responden yang ditunjukkan oleh nilai negatif yang besar dari l_z (biasanya kurang dari -2.00) maka dapat diidentifikasi sebagai pola *aberrant response*.

2. *Outlier*

Outlier adalah suatu data yang dianggap tidak biasa atau mustahil jika dibandingkan dengan keseluruhan data set. Penelitian ini menggunakan data set yang bersifat *univariate*, sehingga digunakan *Modified Z-score* sebagai pendeteksi outlier. *Modified Z-score* menggunakan median dan median dari deviasi mutlak median (MAD) yang dapat dilihat pada persamaan (3) dan persamaan (4). Pada penelitian ini, nilai M_i yang lebih besar daripada 2.0 ditandai berpotensi sebagai outlier.

D. Penetapan Populasi dan Sampel

Penelitian ini menggunakan studi simulasi dan real data. Pada studi simulasi, terdapat dua jenis kondisi jumlah item, yaitu item yang berjumlah 20 dan 40. Setiap kondisi tersebut berisikan 450 responden yang terdistribusi normal. Setelah itu, kedua kondisi akan dikontaminasi dengan 50 responden ber-parameter ektrim. 50 responden ber-parameter ektrim tersebut diharapkan berperan sebagai *aberrant response* yang diidentifikasi dengan menggunakan *person-fit* dan *outlier* yang diidentifikasi dengan *modified Z-score*. Pada *real data*, digunakan 500 subyek yang diambil dari SAPA-Project yang dilaksanakan pada tanggal 08 Desember 2013 hingga 26 Juli 2014.

E. Tahap Pelaksanaan Penelitian

Penelitian ini dilaksanakan dalam beberapa tahap, dimana tahap tersebut terbagi menjadi dua, yaitu tahap simulasi data dan tahap aplikasi real data. Berikut adalah penjelasan dari tiap-tiap tahapan.

1. Simulasi data

Untuk melakukan simulasi data, penelitian ini menggunakan aplikasi R studio versi 1.1.453. R studio merupakan lingkungan pengembangan terintegrasi untuk R. Sedangkan R sendiri merupakan suatu bahasa pemrograman yang banyak digunakan oleh peneliti, karena gratis, fleksibel, dan mampu melakukan berbagai analisis statistik dan visualisasi data (Bulut & Sunbul, Monte Carlo Simulation Studies in Item Response Theory with the R Programming Language, 2017). Berikut adalah tahapan simulasi data:

- a. Peneliti menentukan beberapa faktor simulasi yang diharapkan mempengaruhi operasi prosedur statistik tertentu. Dalam penelitian ini, peneliti memiliki tujuan untuk meneliti *aberrant response* sebagai penyebab kemunculan *outlier*.
- b. Peneliti menentukan jenis desain eksperimen, yakni dengan 2PL atau Graded Response Model.
- c. Beberapa set data dihasilkan berdasarkan faktor simulasi dan kebutuhan penelitian. Penelitian ini menggunakan dua kondisi jumlah item, yaitu 20 item

dan 40 item dengan masing-masing berisi 450 responden. Kedua kondisi digenerasi dengan parameter yang sama, yaitu: *item discrimination*, $a \sim \ln N(0.2, 0.2)$; *item difficulty*, $b \sim N(0, 1)$; dan *ability*, $\theta \sim N(0, 1)$.

- d. Peneliti melakukan estimasi parameter. Pada penelitian ini, peneliti membandingkan nilai *true* dan *estimated parameter* dari data yang digenerasi.
- e. Selanjutnya, peneliti memeriksa nilai *Bias* dan *RMSE* dari data yang digenerasi untuk menentukan apakah data bagus untuk digunakan.
- f. Setelah data siap digunakan, peneliti mengkontaminasi data dengan respon yang ekstrim sebanyak 50 responden kepada tiap-tiap kondisi. Respon yang ekstrim memiliki parameter *item discrimination*, $a \sim \ln N(0, 1)$; *item difficulty*, $b \sim N(2, 2)$; dan *ability*, $\theta \sim N(2, 2)$.
- g. Peneliti kemudian melakukan uji person-fit pada data simulasi menggunakan *package PerFit* pada R Studio untuk mendeteksi *aberrant response*.
- h. Peneliti melakukan perhitungan *modified Z-score* pada data simulasi dengan menggunakan R Studio untuk mendeteksi keberadaan *outlier*.
- i. Peneliti kemudian melakukan uji korelasi dengan menggunakan *pearson's product moment* pada hasil uji *person-fit* dan nilai *modified Z-score* pada data simulasi.

2. Real data

Pada bagian ini, akan digunakan data nyata sebagai pengaplikasian hasil yang didapatkan oleh studi simulasi. Untuk tahapan ini juga digunakan aplikasi R studio. Berikut adalah tahapan dari real data:

- a. Menggunakan real data yang berisikan 500 responden tes. Pada real data, responden juga mengisi 40 item soal yang bersifat polytomous dengan 6 pilihan jawaban.
- b. Peneliti kembali melakukan estimasi parameter pada *real data* untuk menentukan kelayakan data.
- c. *Real data* kemudian diujikan dengan *person-fit* untuk menemukan respon-respon yang dianggap *aberrant*.
- d. Peneliti menghitung nilai *modified Z-score* pada real data untuk mendeteksi keberadaan *outlier*.
- e. Peneliti kemudian melakukan uji korelasi dengan menggunakan *pearson's product moment* pada hasil uji *person-fit* dan hasil *modified Z-score* pada *real data*.

F. Teknik Analisis Data

Penelitian ini menggunakan beberapa teknik untuk menganalisis data secara keseluruhan. Keseluruhan teknik analisis data dilakukan dengan menggunakan software R Studio.

1. *Monte carlo simulation*

Teknik ini digunakan untuk menghasilkan data simulasi yang dibutuhkan dalam penelitian. Monte Carlo Simulation dapat menghasilkan respon secara *random* namun dalam parameter-parameter yang telah ditentukan. Dalam penelitian ini, peneliti membutuhkan 450 responden tes yang berasal dari distribusi normal dan 50 responden tes yang menghasilkan respon ekstrim.

2. *Person-fit*

Uji *person-fit* digunakan untuk mendeteksi keberadaan *aberrant response*. Statistic yang digunakan pada penelitian ini adalah l_z . Pencarian nilai l_z dapat dilakukan dengan persamaan (1) dan kemudian dilanjutkan dengan persamaan (2). Pola respon dari responden yang ditunjukkan oleh nilai negatif yang besar dari l_z (biasanya kurang dari -2.00) maka dapat diidentifikasi sebagai pola *aberrant response*.

3. *Modified Z-score*

Teknik ini digunakan untuk mendeteksi dan menentukan jarak *outlier* pada data set. *Modified Z-score* menggunakan median dan median dari deviasi mutlak median (MAD) yang dapat dilihat pada persamaan (3) dan persamaan (4). Nilai M_i yang lebih besar daripada 2.0 ditandai berpotensi sebagai outlier.

4. *Pearson's product moment*

Pearson's Product Moment digunakan untuk menguji korelasi antara hasil uji *person-fit* dan hasil *modified Z-score*. Untuk mendapatkan nilai korelasi dapat dilihat pada persamaan (5).

$$\rho = \frac{cov(x,y)}{\sigma_x \sigma_y} \quad (5)$$

Dimana $cov(x,y)$ merupakan nilai kovariansi dari nilai x dan nilai y . Sedangkan σ_x merupakan nilai standar deviasi dari x dan σ_y merupakan nilai standar deviasi dari y . Pada penelitian ini variabel x merupakan nilai dari hasil statistik I_z dan variabel y merupakan nilai *modified Z-score* pada data. Apabila nilai korelasi yang muncul lebih besar dari 0, maka menunjukkan korelasi yang positif. Sedangkan jika nilai korelasi kurang dari 0 maka menunjukkan korelasi yang negatif.

BAB IV

HASIL DAN PEMBAHASAN

A. Hasil Penelitian

1. Hasil Simulasi

Sesuai dengan tahapan penelitian, pada tahap pertama simulasi dilakukan generasi data. Generasi data dilakukan dengan *Monte Carlo Simulation* pada *software* R-Studio. Pada studi simulasi penelitian ini, digunakan data set dengan 2 kondisi, dimana kondisi pertama berisikan 20 item dan kondisi kedua berisikan 40 item. Kedua kondisi item memiliki pilihan jawaban sebanyak 6 pilihan. Masing-masing kondisi berisikan 450 responden dengan mean *ability*=0 dan standar deviasi=1. Nilai dari *true parameters* dan *estimated parameter* dari simulasi dapat dilihat pada tabel 2.

Tabel 2.

Nilai true parameter dan estimated parameter dari data simulasi

Parameter	Jumlah Item	α		β_1		β_2		β_3		β_4		β_5	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
True	20	1.247	.253	.989	1.053	.493	1.026	-.007	1.018	-.508	1.025	-1.009	1.048
	40	1.245	.253	1.001	1.048	.503	1.027	.002	1.019	-.497	1.027	-.997	1.051
Estimated	20	1.251	.291	.994	1.074	.486	1.044	-.024	1.035	-.534	1.043	-1.030	1.057
	40	1.232	.285	1.006	1.064	.498	1.040	-.011	1.032	-.519	1.040	-1.015	1.055

Salah satu keuntungan dari melakukan studi simulasi adalah kita dapat mengetahui nilai “*true*” dari parameter data yang kita gunakan. Dimana informasi tersebut dapat memberikan kesempatan untuk melihat *error* yang mempengaruhi

nilai estimasi (Feinberg & Rubright, 2016). Dapat dilihat nilai *true* dan *estimated parameter* pada data set tidak terlalu berbeda jauh, sehingga *estimated parameter* tidak terlalu banyak dipengaruhi oleh *error*.

Setelah menemukan *true parameter* dan *estimated parameter* dari data, tahap selanjutnya adalah melakukan *parameter recovery*. Pada tahap ini, nilai *true parameter* dan *estimated parameter* dibandingkan dengan mencari nilai Bias dan RMSE dari masing masing kondisi. Nilai dari Bias dan RMSE dapat dilihat pada tabel 3.

Tabel 3.
Nilai bias dan RMSE dari tiap tiap kondisi

Jumlah	A		b1		b2		b3		b4		b5	
Item	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
20	.004	.139	.005	.299	-.007	.938	-.017	.788	-.025	.956	-.008	1.319
40	-.013	.134	.005	.301	-.005	.939	-.014	.788	-.022	.954	-.011	1.321

Nilai bias yang positif menandakan nilai *estimated parameter* yang lebih tinggi dibandingkan *true parameter*, sedangkan nilai bias yang negative menandakan sebaliknya. Nilai bias yang baik untuk data simulasi adalah yang mendekati 0, sedangkan nilai RMSE yang baik untuk data simulasi adalah yang mendekati nilai standar deviasi dari *estimated parameter* (Feinberg & Rubright, 2016). Dapat dilihat pada bias bahwa nilai terjauh dari 0 adalah -.025, hal ini menandakan bahwa data simulasi memiliki nilai bias yang bagus untuk digunakan. Perbandingan dari nilai RMSE dan standar deviasi dari *estimated parameter* tidak terpaut jauh, sehingga data simulasi bagus untuk digunakan.

Setelah memiliki data simulasi yang bagus, dilanjutkan dengan mengkontaminasi data dengan respon yang ekstrim. Hal ini dilakukan sesuai dengan tujuan penelitian, yaitu untuk menguji *aberrant response* sebagai penyebab kemunculan *outlier*. Respon ekstrim yang digenerasi diharapkan dapat dianggap sebagai *aberrant response* maupun *outlier*. Respon esktrim dimasukkan pada masing-masing kondisi, dengan *true parameter* yang dapat dilihat pada tabel 4.

Tabel 4.
Parameter simulasi respon yang esktrim

Jumlah Responden	α		β		θ	
	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>
50	1	0	2	2	2	2

Respon ekstrim yang digenerasi kemudian dimasukkan ke dalam data normal yang sudah ada. Respon ekstrim yang dimasukkan berjumlah 50 responden, yaitu sebanyak 10% dari keseluruhan data, sehingga masing-masing kondisi data kini berjumlah 500 responden. Jumlah 500 responden merupakan jumlah yang ideal dalam studi simulasi agar tercipta parameter yang lebih struktural (Reise & Yu, 1990). Selanjutnya, *data set* diujikan dengan person-fit model I_z untuk mendeteksi *aberrant response* pada data. Semakin rendah nilai I_z maka semakin memungkinkan untuk dianggap sebagai *aberrant response*. Hasil dari uji person-fit dapat dilihat pada tabel 5.

Tabel 5.
Hasil uji person-fit pada data set

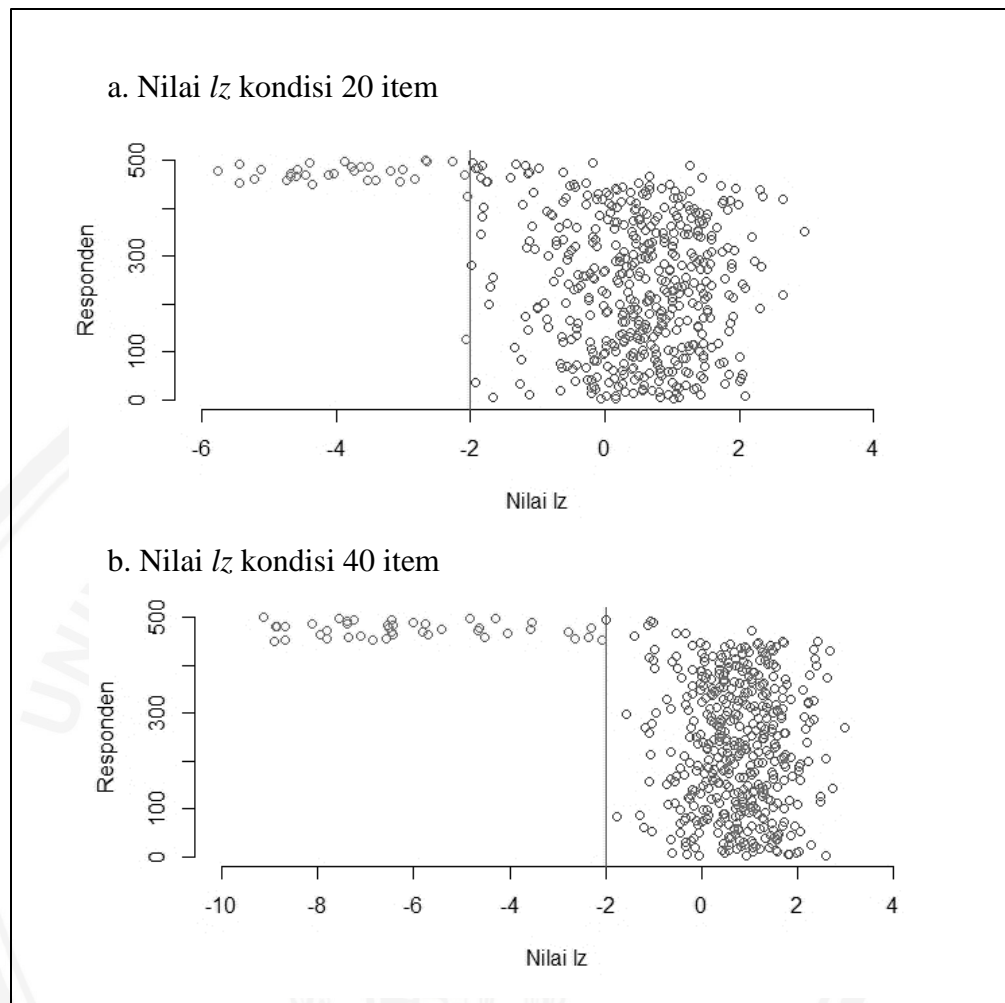
Jumlah item	Nilai l_z Ter-tinggi	Nilai l_z Ter-rendah	Rata-rata nilai l_z	Jumlah Aberrant Response
20	2.964	-5.764	.220	32
40	2.987	-9.121	.160	42

Pada kondisi pertama, ditemukan jumlah *aberrant response* sebanyak 32 atau 6.4% dari keseluruhan data. Dari jumlah *aberrant response* tersebut, 30 respon merupakan hasil dari respon ekstrim yang digenerasi. Sedangkan pada kondisi kedua, ditemukan jumlah *aberrant response* sebanyak 42 respon atau 8.4% dari keseluruhan data dan seluruhnya berasal dari respon esktrim yang digenerasi. Plot yang menggambarkan hasil dari nilai l_z pada data dapat dilihat pada Gambar 2.

Data set kemudian ditentukan nilai *modified z-score*-nya agar dapat menentukan data yang dianggap sebagai outlier. Semakin tinggi nilai *modified z-score* maka semakin memungkinkan untuk dianggap sebagai data *oulter*. Hasil dari penghitungan *modified z-score* dapat dilihat pada tabel 6.

Tabel 6.
Hasil penghitungan modified z-score pada data set

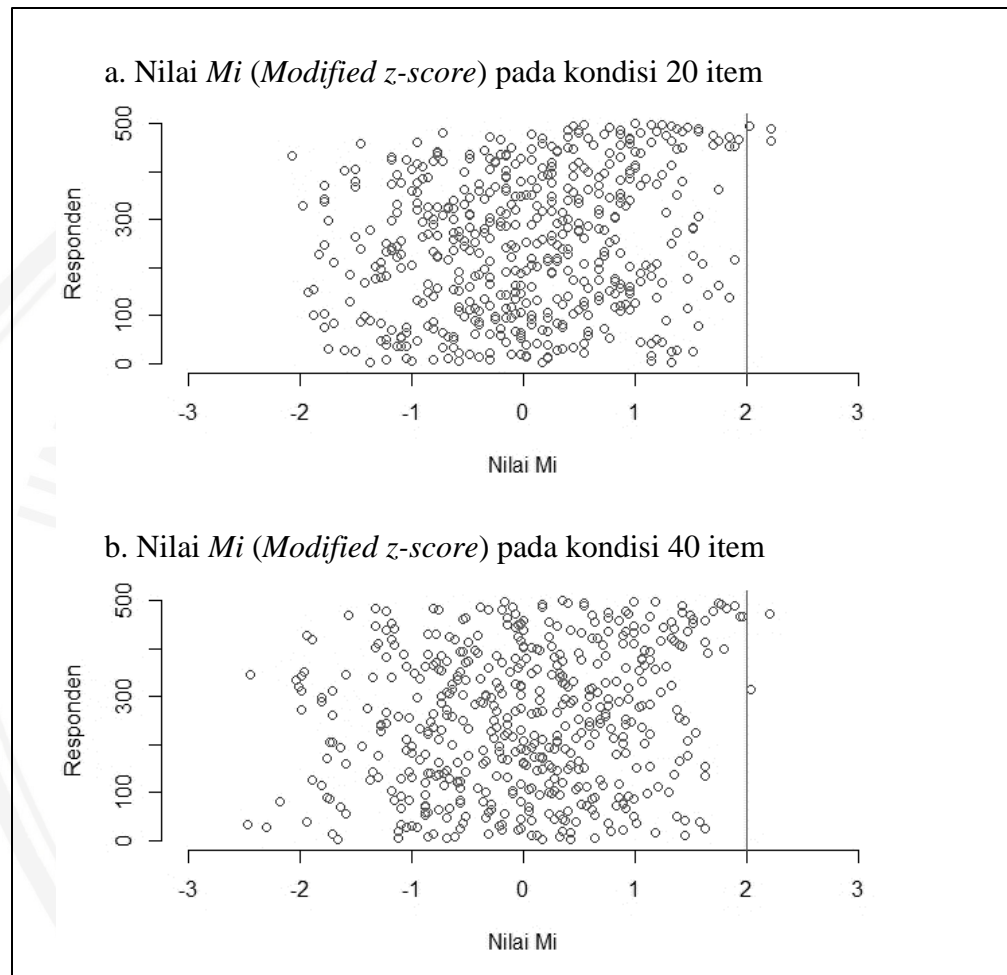
Jumlah item	Nilai M_i Ter-tinggi	Nilai M_i Ter-rendah	Rata-rata nilai M_i	Jumlah Outlier
20	2.209	-2.070	.007	3
40	2.207	-2.477	-.015	2



Gambar 2. Hasil penghitungan nilai I_z dari masing-masing kondisi item

Pada kondisi pertama, ditemukan jumlah *outlier* sebanyak 3 respon atau 0.6% dari keseluruhan data. Dari jumlah outlier tersebut, seluruhnya terdeteksi dari respon ekstrim yang digenerasi. Sedangkan pada kondisi kedua, ditemukan jumlah outlier sebanyak 2 respon atau 0.4% dari keseluruhan data. Dari jumlah outlier yang ditemukan tersebut, hanya 1 respon yang merupakan hasil dari respon esktrim yang

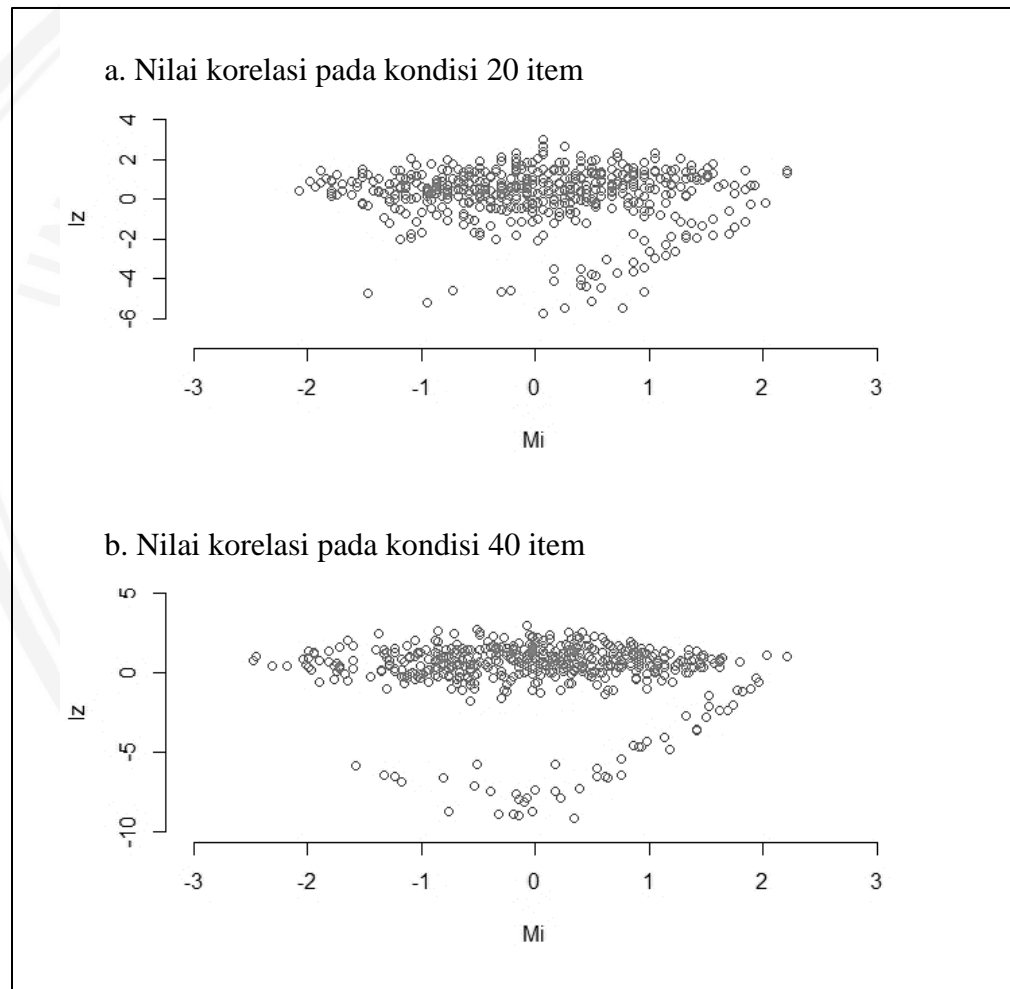
digenerasi. Plot yang menggambarkan hasil dari penghitungan *modified z-score* dapat dilihat pada gambar 3.



Gambar 3. Nilai Mi (*modified z-score*) dari masing-masing kondisi item

Pada kondisi pertama didapatkan 32 *aberrant response*, dan dari jumlah *aberrant response* tersebut tidak ada yang juga terdeteksi sebagai *outlier*. Selanjutnya pada kondisi kedua, didapatkan 42 *aberrant response*, dan dari jumlah tersebut tidak ada respon yang juga terdeteksi sebagai *outlier*. Untuk lebih jelasnya,

nilai l_z dan nilai *modified z-score* dikorelasikan dengan menggunakan *pearson's product moment*. Pada kondisi pertama dan kedua didapatkan nilai korelasi sebesar $-.048$ dan $-.044$. Kedua hasil tersebut dapat digambarkan sebagai tidak ada korelasi menurut Mukaka (2012). Plot korelasi dari kondisi pertama dan kedua dapat dilihat pada gambar 4.



Gambar 4. Hasil korelasi dari masing-masing kondisi item

2. Hasil Real Data

Penelitian ini menggunakan *real data* yang bersumber dari SAPA-Project yang diambil pada tanggal 08 Desember 2013 hingga 26 Juli 2014. SAPA-Project menyebarkan kuesioner kepada kurang lebih 24.000 subyek, namun untuk penelitian ini hanya digunakan 500 subyek yang sudah terpilih berdasarkan jumlah item yang terisi, untuk menghindari kesulitan saat menentukan *outlier*. Data yang digunakan merupakan hasil dari skala PS atau *Plasticity/Stability* yang disusun oleh DeYoung (2010). Skala PS memiliki 40 item pertanyaan dan 6 pilihan jawaban. Berikut adalah nilai estimated parameter yang didapatkan dari *real data*.

Tabel 7.

Nilai item discrimination dan item location dari tiap item real data

Item	β_1	β_2	β_3	β_4	β_5	α
1	-5.855	-4.285	-3.920	-1.015	1.489	.072
2	-1.323	-.236	.406	-2.200	-.762	-.114
3	-1.422	1.956	-.290	3.065	4.066	.208
4	-3.524	-3.236	-3.421	.334	3.879	.071
5	-3.185	-4.876	-8.413	-.104	1.609	.135
6	-3.344	-1.633	.117	-1.698	-1.825	-.087
7	-3.405	-4.542	-5.779	-2.124	.526	.097
8	-2.434	-2.186	-7.590	-3.617	5.939	.118
9	-3.766	-3.793	-6.892	-5.448	-2.837	.179
10	-2.014	-3.499	-7.466	-1.560	1.886	.126
11	-5.506	-2.310	-7.651	-7.288	.236	.094
12	-4.463	-4.825	-5.080	-1.268	2.272	.049
13	-4.704	-4.848	-5.360	2.229	4.434	.072
14	-.984	.334	-1.006	.490	1.044	.389
15	-.427	1.217	-.171	.407	4.344	.290

(lanjut)

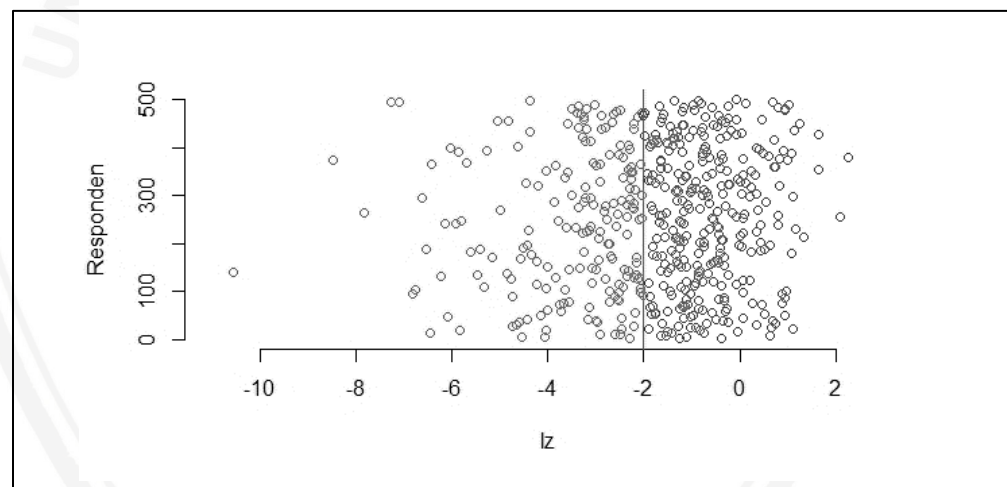
Item	β_1	β_2	β_3	β_4	β_5	α
16	-5.306	-3.550	-9.675	.668	.503	.181
17	-6.156	-5.941	-5.492	.144	.967	.189
18	-5.319	-4.348	-3.181	-.084	2.243	.114
19	-6.303	-8.526	-4.104	-2.921	0	.186
20	-1.015	.062	.017	.996	1.099	.422
21	-1.414	.362	-1.494	1.765	.976	.513
22	-5.413	-4.192	-8.134	-2.713	2.545	.136
23	-.637	.621	-.115	1.702	1.920	.823
24	-1.079	-.318	-.802	.313	.173	.800
25	-.382	.301	.051	.987	1.701	1.418
26	-.625	.142	-.721	.819	.540	.905
27	-1.927	-.484	-1.191	.001	1.129	.847
28	-.257	.260	.040	1.507	2.322	.917
29	-2.693	-.309	-1.337	.249	.700	.797
30	-1.213	-.169	-.394	.437	1.199	1.407
31	-.199	-.269	-.224	1.125	2.568	.618
32	-.026	.474	.923	1.343	2.837	.606
33	-1.074	.192	.269	1.270	2.188	.927
34	-.458	.430	-.015	.883	1.428	1.113
35	-2.277	-.790	-1.260	-.279	1.934	.607
36	-.700	.081	.149	3.004	4.070	.249
37	-1.320	1.807	-4.030	4.801	5.004	.183
38	.808	1.027	-.167	.365	-1.660	-.516
39	-1.662	.539	-1.267	1.198	1.673	.622
40	-.239	.176	-.225	1.144	1.031	1.195

Sesuai dengan klasifikasi nilai *item discrimination* yang disusun oleh Baker (2001), item berentang dari nilai rendah (0.01 - 0.34) hingga tinggi (1.35 – 1.69). Item dengan nilai *discrimination* yang rendah digambarkan semakin sulit untuk membedakan tingkat kemampuan dari responden. Sedangkan untuk *item difficulty*, menyesuaikan dengan klasifikasi dari penelitian Hasmy (2014), *item real data* memiliki tingkat *difficulty* yang berkisar antara *very low* hingga *very high*.

Setelah melakukan estimasi parameter, real data kemudian diujikan dengan *person-fit* model lz . Setelah diuji, terdapat 188 responden atau sebanyak 37.6% dari keseluruhan data yang terdeteksi sebagai *aberrant response*. Informasi mengenai hasil uji person-fit pada real data dapat dilihat pada tabel 8. Plot hasil perhitungan nilai lz pada real data juga dapat dilihat pada gambar 5.

Tabel 8.
Informasi nilai lz dari real data

Jumlah item	Nilai lz Ter-tinggi	Nilai lz Ter-rendah	Rata-rata Nilai lz	Jumlah Aberrant Response
500	2.259	-10.566	-1.726	188



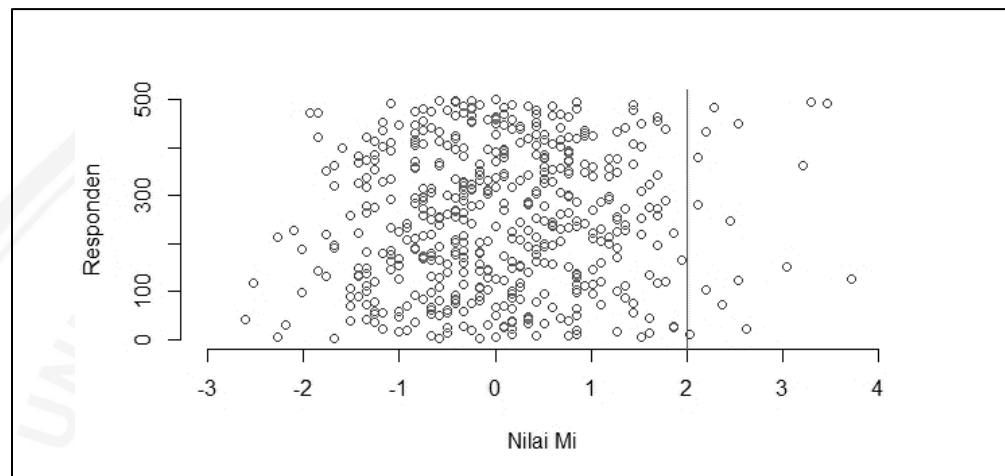
Gambar 5. Hasil penghitungan lz pada real data

Real data kemudian dihitung nilai *modified z-score*nya agar dapat ditentukan respon yang dianggap outlier. Dari 500 respon *real data*, terdeteksi 16 respon atau sebanyak 3.2% merupakan *outlier*. Informasi dari penghitungan nilai *modified z-score* dapat dilihat pada tabel 9 dan gambar 6.

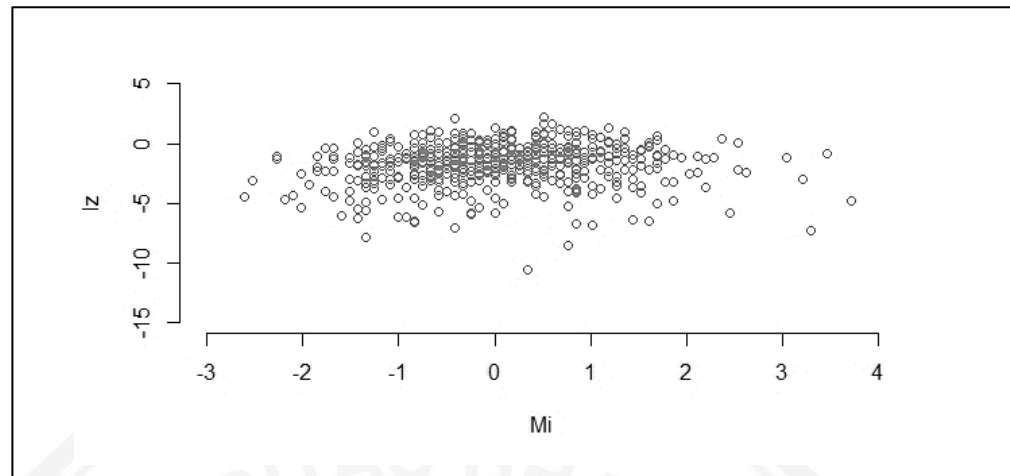
Tabel 9.

Informasi nilai *modified z-score* dari real data

Jumlah item	Nilai Mi Ter-tinggi	Nilai Mi Ter-rendah	Rata-rata Nilai Mi	Jumlah Outlier
500	3.709	-2.613	.025	16

Gambar 6. Hasil penghitungan nilai *Mi* (*Modified z-score*) pada real data

Seperti yang sebelumnya diketahui, pada *real data* ditemukan 188 responden yang dianggap sebagai *aberrant response*. Dari seluruh jumlah respon tersebut, 9 respon diantaranya juga terdeteksi sebagai outlier. Agar lebih jelasnya, nilai I_z dan nilai *modified z-score* dari *real data* dikorelasikan dengan menggunakan *pearson's product moment*. Didapatkan bahwa kedua nilai tersebut memiliki nilai korelasi sebesar .095, dimana menurut Mukaka (2012) nilai tersebut digambarkan sebagai tidak ada korelasi. Gambaran plot dari hasil korelasi kedua nilai dapat dilihat pada gambar 8.



Gambar 7. Hasil korelasi dari real data

B. Pembahasan

Penelitian ini menggunakan studi simulasi dan real data untuk memeriksa *aberrant response* sebagai penyebab munculnya *outlier*. Pada studi simulasi, diketahui bahwa selain data ekstrim yang terdeteksi sebagai *aberrant response* ataupun *outlier*, data yang terdistribusi normal pun juga terkadang terdeteksi. Hal ini disebabkan karena dalam data distribusi normal pun turut memiliki nilai-nilai yang ekstrim, meskipun nilai tersebut hanya sebagian kecil dari keseluruhan data.

Digunakan dua jenis kondisi item pada studi simulasi, yaitu kondisi 20 item dan 40 item. Setelah melakukan uji *person-fit* dengan model *Iz* pada kedua kondisi data, didapatkan beberapa temuan. Pada kondisi 20 item, 6.4% dari keseluruhan data yang digenerasi terdeteksi sebagai *aberrant response*. Sedangkan pada kondisi 40 item, 8.4% dari keseluruhan data yang digenerasi terdeteksi

sebagai *aberrant response*. Hal ini sesuai dengan hasil yang ditemukan oleh penelitian Conijn, Emons, dan Sijtsma (2013) dimana model I_z memiliki kekuatan yang bagus pada tes panjang (> 60 item) dan tes sedang (30 item). Dalam penelitian Spoden (2014), panjang tes 20 item diklasifikasi sebagai *small size lengths* dan panjang tes 40 item diklasifikasi sebagai *medium size lengths*. Kedua jenis panjang tes tersebut dianggap sebagai jenis panjang tes yang paling “menantang” bagi person-fit, karena sesuai dengan penelitian sebelumnya (Drasgow, Levine, & Williams, 1985), person-fit bekerja secara cukup untuk tes yang lebih panjang. Menurut Emons, Glas, Meijer, & Sijtsma (2003), solusi untuk menghadapi kekurangan dari *person-fit* ini hanyalah dengan menambahkan item pada tes atau hanya menggunakan tes dengan kualitas tinggi.

Selanjutnya, kedua kondisi dihitung nilai *modified z-score* dari masing-masing respondennya. Iglewicz dan Hoaglin (1993) menyarankan untuk menandai nilai *modified z-score* di atas 3.5 sebagai *outlier*, namun pada penelitian ini menggunakan nilai *cut-off* yang lebih rendah, yaitu 2. Pada kondisi pertama hanya terdeteksi 3 *outlier* dari keseluruhan data, sedangkan pada kondisi kedua hanya terdeteksi 2 *outlier* dari keseluruhan data. Dari 5 *outlier* tersebut, hanya 1 *outlier* yang berasal dari data ektrim yang digenerasi. Sedangkan pada penelitian Rustum & Adeloye (2007) membuktikan bahwa pendeteksian outlier dengan menggunakan *modified z-score* lebih banyak mendapatkan hasil dibandingkan metode lain (z -

score dan *visual inspection*). Hal ini mungkin disebabkan oleh kurang ekstrim-nya data respon yang dibuat pada data simulasi.

Setelah menguji *person-fit* dan *modified z-score*, hasil dari pengujian tersebut dikorelasikan. Pada kondisi pertama, terdapat 32 *aberrant response* yang terdeteksi dengan *person-fit* sedangkan pada kondisi kedua terdapat 42 *aberrant response*. Namun dari keseluruhan temuan *aberrant response* tersebut, tidak ada yang juga terdeteksi sebagai *outlier*. Temuan tersebut didukung dengan hasil korelasi yang kecil pada kedua kondisi, yaitu -.048 dan -.044, dimana digambarkan sebagai tidak ada korelasi (Mukaka, 2012). Hasil ini hampir sama dengan hasil penelitian dari Widhiarso dan Sumintono (2016), di mana responden yang secara stabil menghasilkan *aberrant response* tidak selalu dianggap sebagai *outlier* pula. Hal ini mungkin disebabkan karena *person-fit* mendeteksi melalui pola respon namun outlier hanya mendeteksi melalui hasil *total score*.

Real data yang digunakan di penelitian ini juga melalui tahap yang sama dengan data simulasi. Dimana pada tahap awal, real data dideteksi dengan menggunakan *person-fit* model *lz*. Berdasarkan uji tersebut, didapatkan 188 responden atau 37.6% terdeteksi sebagai *aberrant response* dengan nilai terendah sebesar -10.556. Hasil yang didapat ini cukup berbeda dengan studi simulasi. Banyaknya responden yang memiliki *aberrant response* pada *real data* kemungkinan disebabkan oleh motivasi pengisi tes, dimana hal ini sesuai dengan penelitian Schmitt, Chan, Sacco, McFarland, dan Jennings (1999) yang

menunjukkan bahwa nilai l_z yang negatif dan terlalu rendah mungkin manifestasi dari motivasi pengerjaan tes yang rendah.

Setelah diujikan dengan *person-fit*, *real data* kemudian dihitung nilai *modified z-score*nya untuk menentukan data *outlier*. Dari cut-off setinggi 2.0, didapatkan 16 respon atau sebanyak 3.2% dari keseluruhan data yang dianggap sebagai outlier. Diketahui bahwa terdapat lebih banyak *outlier* yang terdeteksi pada *real data* dibandingkan data simulasi, hal ini mungkin disebabkan karena data yang *error*, *misreporting* yang disengaja, kesalahan pengambilan sampel, kegagalan standarisasi, dan/atau asumsi distribusi yang salah (Osborne & Overbay, 2004) seperti yang sudah dibahas pada BAB II. Namun sedikitnya outlier yang terdeteksi dapat disebabkan jenis kepribadian *plasticity* dan *stability* yang tersebar secara rata pada sampel yang diteliti. Sehingga, data yang seharusnya dianggap sebagai outlier memiliki nilai *modified z-score* dibawah 2.0.

Kemudian, hasil pengujian *person-fit* dan *modified z-score* dari real data dikorelasikan. Dari 188 responden yang terdeteksi sebagai *aberrant response*, hanya 18 atau 9.57% yang juga terdeteksi sebagai *outlier*. Hasil ini didukung dengan nilai korelasi yang rendah, yaitu .095, dimana menurut Mukaka (2012) nilai tersebut tidak dianggap berkorelasi. Berdasarkan hasil korelasi tersebut, dapat disimpulkan bahwa hasil *real data* sesuai dengan hasil dari data simulasi. Hal ini berarti sesuai dengan prinsip dari studi simulasi Monte Carlo, yaitu prinsip *authenticity*, dimana studi simulasi dapat mencerminkan kondisi nyata (Bulut &

Sunbul, 2017). Namun hasil penelitian ini hanya berlaku pada hasil tes skala PS atau *Plasticity/Stability* yang merupakan skala *personality low-stake test*. Oleh karena itu, dianjurkan untuk meneliti ulang dengan jenis skala lain.

C. Keterbatasan Penelitian

Penelitian ini bertujuan untuk menguji *aberrant response* sebagai penyebab kemunculan dari *outlier*. Namun pada data simulasi, respon yang digenerasi sebagai ekstrim tidak terdeteksi sebagai *outlier*. Selain itu jumlah item yang terlalu sedikit pada data simulasi menyebabkan *person-fit* model *Iz* kurang bekerja secara maksimal pada penelitian ini.

BAB V

PENUTUP

A. Kesimpulan

Penelitian ini bertujuan untuk melihat *aberrant response* sebagai penyebab dari kemunculan *outlier*. Berdasarkan hasil yang didapatkan dari studi simulasi dan *real data*, tidak terdapat korelasi antara kemunculan *aberrant response* dan kemunculan *outlier* pada data. Sehingga, responden yang menghasilkan pola respon yang *aberrant* pada tes tidak selalu menghasilkan suatu nilai akhir yang ekstrim. Selain itu penelitian ini juga menemukan meningkatnya kinerja *person-fit model* I_z apabila ada peningkatan jumlah item pada data set. Sehingga semakin banyak item pada data set, maka *person-fit model* I_z akan bekerja lebih baik dalam mendeteksi *aberrant response*.

B. Saran

Setelah melakukan penelitian dan penarikan kesimpulan, berikut adalah saran berdasarkan penelitian yang sudah dilakukan:

1. Ada baiknya untuk melakukan penelitian ulang dengan topik yang sama namun dengan jumlah item yang lebih banyak (>60).
2. Untuk penelitian selanjutnya, disarankan untuk mencoba uji *person-fit* dan *outlier* pada jenis skala tes yang lain.

3. Disarankan untuk melakukan studi simulasi mengenai jenis parameter yang dapat menghasilkan jenis data *outlier* dan/atau *aberrant response*.
4. Untuk pengguna tes berjenis *low-stake test*, disarankan untuk mendeteksi dan menghilangkan *aberrant response* agar hasil lebih valid. Namun tidak disarankan untuk menghapus nilai *outlier* secara sembarangan karena nilai tersebut dapat disebabkan oleh banyak faktor.



DAFTAR PUSTAKA

- Anscombe, F. J. (1960). Rejection of outliers. *Technometrics*, 2(2), 123-146.
- Anusha, A., & Rao, I. S. (2018). A Study on Outlier Detection for Temporal Data. *International Journal of Engineering Science and Computing*, 8(3).
- Artner, R. (2016). A simulation study of person-fit in the Rasch model. *Psychological Test and Assessment Modeling*, 58(3), 531-563.
- Baker, F. B. (2001). *The Basics of Item Response Theory*. USA: ERIC Clearinghouse on Assessment and Evaluation.
- Batenburg, T. A., Hendriks, A. A., & Kuyper, H. (2012). Atypical Testing Behaviour on a Low Stakes Intelligence Test . *Groningen Institute for Educational Research (GION)*.
- Bethlehem, J. (2009). *Applied survey methods*. Hoboken: John Wiley & Sons, Inc.
- Bulut, O., & Sunbul, Ö. (2017). Monte Carlo Simulation Studies in Item Response Theory with the R Programming Language. *Journal of Measurement and Evaluation in Education and Psychology*, 8(3), 266-287.
- Bulut, O., & Sunbul, Ö. (2017). Monte Carlo Simulation Studies in Item Response Theory with the R Programming Language . *Journal of Measurement and Evaluation in Education and Psychology*, 266-287.
- Cole, J. S., & Osterlind, S. J. (2008). Investigating Differences Between Low- and High-Stakes Test Performance on a General Education Exam. *The journal of general education*, 57(2).
- Condon, D. M., & Revelle, W. (2015). Selected personality data from the SAPA-Project: 08Dec2013 to 26Jul2014. <https://doi.org/10.7910/DVN/SD7SVE> Harvard Dataverse, V4, UNF:6:9B60JNIFXXKXmyLhABjoUA== [fileUNF] .
- Conijn, J. M., Emons, W. H., & Sijtsma, K. (2013). Statistic lz-Based Person-Fit Methods for Noncognitive Multiscale Measures. *Applied Psychological Measurement*, 38(2), 122-136.

- Drasgow, F., Levine, M. V., & Williams, E. A. (1985). Appropriateness measurement with polytomous item response models and standardized indices. *British Journal of Mathematical and Statistical Psychology*, 67-86.
- Emons, W. H., Glas, C. A., Meijer, R. R., & Sijtsma, K. (2003). Person Fit in Order-Restricted Latent Class Models. *APPLIED PSYCHOLOGICAL MEASUREMENT*, 27(6).
- Feinberg, R. A., & Rubright, J. D. (2016). Conducting Simulation Studies in Psychometrics. *Educational Measurement: Issues and Practice*, 35(2), 36-49.
- Fel, J. M., Castaneda, R., Tiemensma, J., & Depaoli, S. (2017). Using Person Fit Statistics to Detect Outliers in Survey Research. *Frontiers in Psychology*, 8(863).
- Hasmy, A. (2014). Compare Unidimensional & Multidimensional Rasch Model for Test with Multidimensional Construct and Items Local Dependence. *Journal of Education and Learning*, 8(3), 187-194.
- Hirsh, J. B., DeYoung, C. G., & Peterson, J. B. (2009). Metatraits of the Big Five Differentially Predict Engagement and Restraint of Behavior. *Journal of Personality*, 77(4).
- Hodge, V. J., & Austin, J. (2004). A Survey of Outlier Detection Methodologies. *Artificial Intelligence Review*, 85-126.
- Iglewicz, B., & Hoaglin, D. (1993). Volume 16: How to Detect and Handle Outliers. *The ASQC Basic References in Quality Control: Statistical Techniques*.
- Jiang, S., Wang, C., & Weiss, D. J. (2016). Sample Size Requirements for Estimation of Item Parameters in the Multidimensional Graded Response Model. *Frontiers in Psychology*, 7(109).
- Karabatsos, G. (2003). Comparing the Aberrant Response Detection Performance of Thirty-Six Person-Fit Statistics. *Applied Measurement in Education*, 16(4), 277-298.
- Li, M. F., & Olejnik, S. (1997). The power of Rasch person-fit statistics in detecting unusual response patterns. *Applied Psychological Measurement*, 215-231.
- Luecht, R., & Ackerman, T. A. (2018). A Technical Note on IRT Simulation Studies: Dealing With Truth, Estimates, Observed Data, and Residuals. *Educational Measurement: Issues and Practice*, 1-12.

- Meijer, R. R. (1996). Person-Fit Research: An Introduction. *APPLIED MEASUREMENT IN EDUCATION*, 9(1), 3-8.
- Meijer, R. R. (2003). Diagnosing Item Score Patterns on a Test Using Item Response Theory-Based Person-Fit Statistics. *Psychological Methods*, 8(1), 72-87.
- Meijer, R. R., & Sijtsma, K. (2001). Methodology Review: Evaluating Person Fit. *Applied Psychological Measurement*, 25(2), 107-135.
- Meijer, R. R., Niessen, A. S., & Tendeiro, J. N. (2016). A Practical Guide to Check the Consistency of Item Response Patterns in Clinical Research Through Person-Fit Statistics: Examples and a Computer Program. *Assessment*, 23(1), 52-62.
- Mukaka, M. (2012). A guide to appropriate use of Correlation coefficient in medical research. *Malawi Med J*, 24(3), 69-71.
- Nader, I. W., Tran, U. S., & Voracek, M. (2015). Effects of Initial Values and Convergence Criterion in the Two-Parameter Logistic Model When Estimating the Latent Distribution in BILOG-MG3. *PLoS One*, 10(10).
- Osborne, J. W., & Overbay, A. (2004). The Power of Outliers (and Why Researchers Should Always Check for Them). *Practical Assessment, Research, and Evaluation*, 9(6).
- Penny, K. I. (1996). Appropriate Critical Values When Testing for a Single Multivariate Outlier by Using the Mahalanobis Distance. *Applied Statistics*, 45(1), 73-81.
- Pllana, S. (2000, Agustus). *History of Monte Carlo method*. Retrieved from <http://www.geocities.com/CollegePark/Quad/2435/index.html>
- Reise, S. P., & Yu, J. (1990). Parameter Recovery in the Graded Response Model Using MULTILOG. *Journal of Educational Measurement*, 27(2), 133-144.
- Rustum, R., & Adeloje, A. J. (2007). Replacing Outliers and Missing Values from Activated Sludge Data Using Kohonen Self-Organizing Map. *JOURNAL OF ENVIRONMENTAL ENGINEERING*, 909-916.
- Schmitt, N., Chan, D., Sacco, J. M., McFarland, L. A., & Jennings, D. (1999). Correlates of Person Fit and Effect of Person Fit on Test Validity. *Applied Psychological Measurement*, 41-53.
- Schmitt, N., Chan, D., Sacco, J. M., McFarland, L. A., & Jennings, D. (1999). Correlates of person fit and effect of person fit on test validity. *Applied Psychological Measurement*, 41-53.

- Seo, S. (2006). A Review and Comparison of Methods for Detecting Outliers in Univariate Data Sets (Thesis). *University of Pittsburgh*.
- Spoden, C. (2014). *Person Fit Analysis with Simulation-based Methods*. Essen: Universität Duisburg.
- Widhiarso, W., & Sumintono, B. (2016). Examining response aberrance as a cause of outliers in statistical analysis. *Personality and Individual Differences*, 11-15.

